

## *Contrastes de correlación serial en las perturbaciones de un modelo econométrico*

---

### I. INTRODUCCIÓN

En la literatura econométrica, aparece con frecuencia citado el llamado modelo lineal general

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + u_t \quad [1]$$

con el que se intenta poner de relieve la relación existente entre las variables conformadoras de un determinado proceso económico. En dicho modelo, los elementos característicos son, por una parte, las variables, de las que  $Y_t$  es la endógena o «a explicar»,  $X_{1t}, X_{2t}, \dots, X_{kt}$  son las predeterminadas, o «explicativas», y  $u_t$  es un término «residual» o de perturbación; y por otra los parámetros  $\beta_0, \beta_1, \dots, \beta_k$ , que sirven para ponderar la incidencia relativa de cada variable en el proceso, y sobre cuya magnitud se interesa la Econometría.

Todas las anteriores variables, a excepción de  $u_t$ , son observables, a poco que se instrumenten los mecanismos válidos para ello. Tal observación puede venir referida, en unos casos, a la sucesión continuada de valores que tales variables alcanzan en períodos consecutivos de tiempo, dando lugar a los llamados modelos temporales; y en otros, a los valores que adoptan bajo criterios no temporales, apareciendo en tal caso los modelos de corte transversal, también conocidos con el nombre de modelos atemporales.

La variable  $u_t$  no posee un contenido tan estrictamente relevante en el modelo, y su presencia obedece a un conjunto de razones, que podríamos denominar «accidentales», de entre las que destacan dos por su importancia. En primer lugar se introduce para compensar la falta de posibles variables explicativas, que no se han incluido explícitamente, bien por desconocimiento de su existencia, bien en razón de su escasa significación separada. En segundo lugar está el que el modelo [1] se supone lineal, lo cual en multitud de ocasiones no es más que una aproximación a la realidad. Ambas son suficientes para comprender que el comportamiento de  $u_t$  no es aprensible, resultando que

en consecuencia, y a diferencia de lo que ocurre con las restantes variables, ésta no puede observarse.

Ante la situación anterior, la solución que tradicionalmente se ha propuesto para suplir la falta de información sobre el funcionamiento del proceso modelizado debida a la presencia de las  $u_t$ , ha sido admitir para éstas un comportamiento aleatorio en cada punto de observación  $t$ , a partir del cual, y sobre la base de las observaciones disponibles sobre las restantes variables, se ha fundamentado la utilización de un conjunto de métodos estadísticos, que forman parte del contenido específico de la Econometría Teórica, con los que se obtienen estimaciones de los parámetros del modelo, cuya validez depende, en cada caso, del grado de adecuación entre las hipótesis que justifican su aplicación, de una parte, y la distribución que realmente siga la perturbación en relación con el proceso modelizado, de otra.

Hay que apuntar desde este momento que, desgraciadamente, una alta adecuación no es excesivamente frecuente. Ello hay que atribuirlo básicamente al poco tiempo transcurrido desde que la Econometría ha tomado entidad propia como disciplina autónoma, y que por eso, al hallarse en una fase inicial, se ve abocada a la doble tarea de, por una parte, investigar nuevos métodos específicos de aplicación al marco de la Economía, y por otra de justificar su existencia mediante una utilización al nivel que sea; y esto último muchas veces se consigue recurriendo a métodos cuya validez en el contexto de otras ciencias, fundamentalmente de tipo experimental, parece incuestionable, pero que difícilmente pueden aceptarse sin ciertos reparos en el que nos ocupa.

Un ejemplo típico de actuación dentro de la línea anterior lo ofrece la utilización del método de los mínimos cuadrados ordinarios para estimar el modelo [1], al amparo de la presunción de que dicho modelo es de regresión y se satisfacen para las variables que contiene la misma serie de hipótesis que cuando se utiliza en el ámbito de las ciencias de tipo experimental; y con la pretensión de que por tal motivo las estimaciones que se deducen poseerán propiedades análogas a las obtenidas en dichos casos. Entre las hipótesis mencionadas está una que centrará nuestra atención en lo sucesivo. La que establece que la distribución del término de perturbación en cualquier punto  $t$  del dominio  $\tau$  de observación, además de tener esperanza matemática nula, y varianza  $\sigma^2$  constante, debe ser independiente de la distribución en otro punto cualquiera, es decir,

$$\text{Cov } \{u_t, u_{t+h}\} = 0 \quad t, t+h \in \tau \quad h \neq 0 \quad [2]$$

además de

$$E \{u_t\} = 0 \quad \text{y} \quad \text{Var } \{u_t\} = E \{u_t^2\} = \sigma^2$$

hipótesis que suele conocerse como de «ausencia de correlación serial» en el término de perturbación.

En relación con tal hipótesis, la realidad es, como muy bien ha puesto de manifiesto Lozano (1965), que existen diversas razones que inducen a pensar, sobre todo si las observaciones son temporales, que no va a satisfacerse

en el marco del modelo lineal general. Entre ellas destaca poderosamente la que resalta el que como los términos de perturbación representan el efecto conjunto de variables omitidas, y éstas, según indica Orcutt (1948), muestran al ser observadas una correlación serial más o menos acusada, debida al retraso con que reaccionan ante alguno o ante todos los elementos que las determinan, otro tanto ocurre con aquéllos.

Interesa de paso aclarar que la correlación serial en el término de perturbación, aun cuando se presenta con frecuencia en los modelos temporales, no es exclusiva de ellos. En los modelos espaciales puede también ser observada como muestra Nerlove (1963) en relación con la función de costes totales para la electricidad. Sin embargo, no es tan corriente, y presupone la existencia de un orden unidimensional natural en las observaciones semejante al cronológico no siempre fácil de establecer.

La hipótesis que venimos considerando, tal como hemos intentado poner de manifiesto, parece estar basada más que en su plausibilidad económica en la conveniencia estadística de posibilitar la obtención de estimaciones óptimas, mediante el método de los mínimos cuadrados ordinarios (MCO). Resulta en definitiva una muestra del tratamiento indebido de un modelo econométrico al que antes hicimos referencia. Frente a tal situación al econometra se le presentan dos caminos a seguir, que en ningún caso son incompatibles: el primero recurrir a otros métodos de estimación que tomen explícitamente en cuenta la presencia de correlación en las perturbaciones, y otro contrastar la hipótesis de ausencia de correlación con el fin de averiguar el nivel de significación a que resultan admisibles las estimaciones obtenidas. Como quiera que el presente trabajo se orienta en esta segunda dirección, no como vano intento de aclarar la justificación de lo injustificable, como quizá cabría pensar después de lo dicho hasta aquí, sino por una serie de razones que más adelante se detallarán, vale la pena mencionar en primer lugar, aunque sólo sea de pasada, lo más importante realizado dentro de la primera línea.

## II. ESTIMACIÓN DE UN MODELO CON PERTURBACIONES SERIALMENTE CORRELACIONADAS

Consideremos de nuevo el modelo lineal general descrito en el apartado anterior y supongamos que se dispone de una muestra de tamaño  $T$  de observaciones sobre el comportamiento de las variables incluidas en él. Ordenando la información que dicha muestra proporciona en las matrices

$$X = \begin{pmatrix} X'_1 \\ X'_2 \\ \cdots \\ \cdots \\ X'_T \end{pmatrix} = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ X_{1T} & X_{2T} & \cdots & X_{kT} \end{pmatrix} \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ \cdots \\ Y_T \end{pmatrix}$$

en la primera de las cuales cada fila  $X'_t$  está formada por los valores observados en el punto  $t$ , para las  $k$  variables predeterminadas, y en la segunda  $Y_t$  es el valor observado para la endógena en igual punto, es posible escribir de nuevo el modelo, ahora bajo las formas

$$[I] \quad Y = X\beta + u \qquad [II] \quad Y_t = X'_t\beta + u_t \qquad [4]$$

donde

$$\beta' = \|\beta_1 \ \beta_2 \ \dots \ \beta_k\| \quad \text{y} \quad u' = \|u_1 \ u_2 \ \dots \ u_T\|$$

son respectivamente los vectores de parámetros a estimar y de perturbaciones.

Cuando se acepta para el vector  $u$  un comportamiento aleatorio, con parámetros

$$E\{u\} = 0 \qquad E\{uu'\} = \Omega \qquad [5]$$

donde  $\Omega$  es una matriz definida positiva, estamos considerando implícitamente la posibilidad de que exista correlación entre las perturbaciones correspondientes a distintas observaciones. Respondería ésta al caso particular en el que los elementos de  $\Omega$  situados fuera de la diagonal principal son distintos de cero.

Pues bien, bajo las premisas anteriores, una estimación recomendable para  $\beta$  es la proporcionada por el método de los mínimos cuadrados generalizados, debido a Aitken (1934), que adopta la forma:

$$\hat{\beta} = (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} Y) \qquad \text{con} \qquad \text{Var}(\hat{\beta}) = (X' \Omega^{-1} X)^{-1}$$

y posee las propiedades de ser, además de consistente, la de menor varianza entre las lineales e insesgadas, como demuestran entre otros Pena (1965) y Alcaide (1966, pp. 185-190). Este método ha sido reiteradamente utilizado con la finalidad que nos ocupa bajo la hipótesis concreta de que los elementos de  $\Omega$  poseen la estructura:

$$\omega_{t, t \pm k} = \text{Cov}(u_t, u_{t \pm k}) = \sigma^2 \rho^k$$

o, lo que es lo mismo,

$$\rho^k = \frac{\text{Cov}(u_t, u_{t \pm k})}{\sigma^2} = \frac{\text{Cov}(u_t u_{t \pm k})}{D(u_t) D(u_{t \pm k})} \qquad [7]$$

donde  $D$  es el operador Desviación Tipo. Con lo cual,  $\rho^k$  puede ser interpretado como el coeficiente de correlación entre las variables  $u_t$  y  $u_{t \pm k}$ , cumpliendo por ello  $\rho^k \in (-1, 1)$  y será tanto menor cuanto mayor sea la diferencia  $k$  entre



dos puntos de distribución de  $u_t$ , ya que en particular [7] para  $k = 1$  implica que  $\rho \in (-1, 1)$ .

Este planteamiento, según ponen de relieve Goldfeld y Quandt (1972, pp. 181-183), es el correspondiente al siguiente esquema de generación de perturbaciones:

$$u_t = \rho u_{t-1} + \varepsilon_t \quad [8]$$

donde  $\varepsilon_t$  es un ruido blanco distribuido normalmente con parámetros  $(0, \sigma_\varepsilon)$ , y con independencia de  $u_{t+h}$ .<sup>1</sup> Tal esquema expresa un comportamiento para las perturbaciones en cada punto conocido con el nombre de esquema autorregresivo de primer orden, y según muestran Ames y Reiter (1961) traduce una situación habitual en las variables económicas: que el valor que alcanzan en un momento  $t$  es igual a una porción del valor alcanzado en el precedente junto con un efecto aleatorio, y de ahí su aplicabilidad.

Kadiyala (1968) ha demostrado que la estimación [6] bajo la hipótesis particular [7] es precisamente la que se obtiene al aplicar mínimos cuadrados ordinarios al siguiente modelo transformado:

$$Y_t^* = \beta_0 + \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + \dots + \beta_k X_{kt}^* + u_t^* \quad [9]$$

en el que las variables

$$Y_t^* = Y_t - \rho Y_{t-1} \quad X_{it}^* = X_{it} - \rho X_{i(t-1)} \quad u_t^* = u_t - \rho u_{t-1} \quad [10]$$

están definidas para los puntos  $t = 2, 3, 4, \dots, T$ ; y que, por tal motivo, su determinación resultará de la minimización de

$$\sum_2^T e_t^2 = \sum_2^T (Y_t^* - X_t'^* \hat{\beta})^2 \quad [11]$$

la cual se logra para:

$$\hat{\beta} = (\sum_2^T X_t^* X_t^{*'})^{-1} \cdot (\sum_2^T X_t^* Y_t^*) \quad [12]$$

Según todo lo anterior, y a modo de resumen, se puede decir que, dado el modelo [4], si se presupone que la distribución del término de perturbación sigue un esquema autorregresivo de primer orden del tipo [8], lo cual ocurre muchas veces cuando se manejan variables económicas, una estimación óptima de  $\hat{\beta}$  vendrá dada por la expresión [12], como equivalente a la de MCG, en la que los asteriscos indican valores deducidos a partir de las observaciones

1. Para especificaciones más complicadas, véase Wold (1950) y Theil (1971, sección 6.5).

de las variables por medio de [10]. Ahora bien, si se observa atentamente la última relación escrita, se comprueba cómo su aplicación está basada en  $\rho$ , el cual dada su propia naturaleza no es conocido jamás. La única solución que cabe es utilizar en su lugar una estimación suya, pero bien entendido que entonces las estimaciones de  $\beta$  subsiguientes no serán más que aproximaciones a las de MCG, cuya validez dependerá en cada caso de las propiedades asociadas a la  $\hat{\rho}$  que se utilice.

Varias son las soluciones propuestas para estimar  $\rho$  ante la dificultad de hacerlo minimizando directamente la resultante de sustituir [12] en [11] por culpa de los algoritmos que aparecen que son no lineales y complicados de manejar. Entre tales soluciones hay que citar las debidas a Cochrane y Orcutt (1949), Durbin (1960) y Friedman y Meiselman (1963), cuya descripción para el caso particular de un modelo con una sola variable explicativa puede encontrarse en Johnston (1972, pp. 261-264) y Kmenta (1971, pp. 282-292). Todas ellas muestran como característica común el que posibilitan estimaciones de  $\beta$  asintóticamente equivalentes a las de MCG, y que en general, incluso para tamaños muestrales pequeños, son más eficientes que las obtenidas por MCO, salvo que  $|\rho| < 0,30$  en cuyo caso la eficiencia es prácticamente coincidente según prueban por simulación experimental Griliches y Rao (1969) y Kmenta y Gilbert (1968).

Se puede decir, como conclusión de este apartado, que la aplicación del método de los MCG en un modelo econométrico con perturbaciones serialmente correlacionadas no es sencilla, ya que al desconocerse la estructura autorregresiva concreta que opera para tales perturbaciones, son necesarias hipótesis y métodos complementarios para obtener como mucho resultados aproximados. Métodos que por lo demás poseen como característica muy acusada su extremada complejidad operativa, reiterativa en algunos casos como el de Orcutt y Cochrane y sobre cuya eficiencia relativa con los MCO no hay nada escrito de manera definitiva, sobre todo si partiendo de un esquema autorregresivo de primer orden el valor de  $\rho$  es relativamente pequeño.

Por todo ello, parece indispensable disponer de algún mecanismo cuya aplicación nos aporte *a priori* luz suficiente para evitar la utilización de los métodos detallados, caso de ser innecesarios en el sentido de mejorar muy poco las estimaciones por MCO. Tal papel corresponde al contraste estadístico que pasamos a definir.

### III. CONTRASTE DE AUSENCIA DE CORRELACIÓN EN LAS PERTURBACIONES

Dado el modelo lineal general  $Y = X\beta + u$ , en el que el vector  $u$  satisface las hipótesis

$$E\{u\} = 0 \quad E\{uu'\} = \Omega \quad u : N(0, \Omega) \quad [13]$$

un contraste de ausencia de correlación serial es el debido a Von Neumann (1941). Basado en el estadístico

$$\hat{\sigma}^2/S^2 = \frac{\sum_2^T (u_t - u_{t-1})^2/T - 1}{\sum_1^T (u_t - \bar{u})^2/T} \quad [14]$$

conocido como «la razón de la media cuadrática de las diferencias sucesivas respecto a la varianza», sirve para contrastar la hipótesis nula  $H_0: \Omega = \sigma^2 I$  de ausencia de correlación, frente a la alternativa  $H_1$  de que las  $u_t$  estén correlacionadas positivamente. La idea básica que lo fundamenta es la de que el numerador

$$\sum_2^T (u_t - u_{t-1})^2 = \sum_2^T u_t^2 + \sum_2^T u_{t-1}^2 - 2 \sum_2^T u_t u_{t-1}$$

tenderá a ser pequeño bajo la hipótesis  $H_1$ , al ser el sustraendo que aparece en el segundo miembro tanto mayor cuanto mayor sea la dependencia existente entre la distribución de las  $u_t$ .

Durbin y Watson (1950) modifican el estadístico anterior, para el cual asignan la nueva forma

$$d = \frac{\sum_2^T (u_t - u_{t-1})^2}{\sum_1^T (u_t - \bar{u})^2} = \frac{T-1}{T} (\hat{\sigma}^2/S^2) \quad [15]$$

que es como se utiliza en la práctica por su mayor facilidad computacional, y que en lo sucesivo denominaremos «Razón de Von Neumann modificada». Con notación compacta será

$$d = \frac{u' Au}{u' B_T u} \quad [16]$$

en la que

$$A = \begin{vmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{vmatrix} \quad \text{y} \quad B_T = I_T - \frac{1}{T} i i' \quad \text{con} \quad i = \begin{vmatrix} 1 \\ 1 \\ 1 \\ \dots \\ 1 \\ 1 \end{vmatrix}$$

La distribución teórica de  $d$  bajo la hipótesis nula es fácilmente alcanzable al tener en cuenta, tal como se demuestra en el Corolario I del Apéndice, que

$$P\{d \leq d_\alpha/H_0\} = P\left\{\sum_{t=1}^T u_t w_t^2 \leq 0\right\} \quad [17]$$

donde  $\mu_t$  son las raíces características de la matriz  $(A - d_\alpha B_T)$  que pueden determinarse de una vez por todas para diversos valores de  $T$ . Hart (1942) tabuló la distribución anterior<sup>2</sup> para ciertos valores, mediante un método aproximado basado en un ajuste a una distribución beta, demostrando, además, que  $d$  es simétrica con parámetros:

$$E(d) = 2 \quad \text{Var}(d) = 4 \frac{T-2}{T^2-1}$$

Abrahamse y Koerts (1969) han ajustado la distribución con mayor precisión hasta ofrecer la forma que muestra el cuadro 1. Para ello utilizan un procedimiento debido a Imhof (1961) cuyo programa de computación puede encontrarse en Koerts y Abrahamse (1971, pp. 165-174).

Conociendo la distribución de  $d$ , para diversos valores de  $T$ , el contraste a utilizar consistirá en, fijado un nivel de significación  $\alpha$ , averiguar si el valor de  $\hat{d}$  correspondiente a una muestra seleccionada cae dentro de la región crítica de rechazo de  $H_0$ , que será  $\hat{d} \leq d_\alpha$ , o fuera de ella, para rechazar la ausencia de correlación o aceptarla según el caso. Además, la simetría que hemos dicho posee la distribución de  $d$ , permite contrastar la hipótesis de ausencia de correlación negativa, a través de  $4 - d_\alpha$ , el cual, si es menor que  $\hat{d}$  permite rechazar la hipótesis, y si es mayor no, debiendo aceptarla al mismo nivel de significación.

La mayor dificultad con que tropieza la aplicación del contraste anterior está en que las  $u_t$  correspondientes a una realización muestral no son observables, y, por tanto, no pueden utilizarse. Como mucho pueden estimarse, pero en tal caso la distribución del estadístico equivalente

$$d' = \frac{\hat{u}' A \hat{u}}{\hat{u}' B_T \hat{u}} \quad [18]$$

deberá analizarse a la luz de cada estimación  $\hat{u}_t$  que se maneje para averiguar las posibilidades de aplicación que se le pueden asignar. Supongamos que elegimos una estimación de  $u$  satisfaciendo los requisitos:

$$(i) \quad \hat{u} = Ku \quad (ii) \quad E(\hat{u}) = 0 \quad [19]$$

es decir, que sea lineal<sup>3</sup> e insesgada. Según [13] se cumplirá:

$$\text{Var}(\hat{u}) = E(Kuu'K') = K E(uu') K' = K \Omega K' \quad [20]$$

2. En realidad tabuló  $(d^2/S^2)$ , pero sus valores son fácilmente adaptables a  $d$  mediante la transformación  $(T-1)/T \cdot (d^2/S^2)$ .

3. Anderson (1962) justifica la conveniencia de seleccionar un estimador lineal para evitar las complicaciones adicionales derivadas de que en otro caso deberían manejarse los momentos de orden superior al segundo, de los cuales no se posee ninguna información *a priori*.

CUADRO 1

*Valores de la distribución del estadístico de Von Neumann modificado*

T	$\alpha = 1 \%$	2,5 %	5 %	10 %
5	0,5379	0,6645	0,8203	1,0509
6	0,5614	0,7213	0,8902	1,0994
7	0,6140	0,7795	0,9359	1,1407
8	0,6647	0,8247	0,9817	1,1832
9	0,7090	0,8697	1,0245	1,2187
10	0,7517	0,9111	1,0621	1,2511
11	0,7915	0,9486	1,0965	1,2800
12	0,8280	0,9829	1,1276	1,3060
13	0,8619	1,0143	1,1558	1,3295
14	0,8932	1,0432	1,1816	1,3508
15	0,9222	1,0697	1,2053	1,3703
16	0,9492	1,0943	1,2271	1,3882
17	0,9744	1,1171	1,2473	1,4047
18	0,9979	1,1383	1,2660	1,4199
19	1,0199	1,1581	1,2834	1,4341
20	1,0406	1,1766	1,2996	1,4473
21	1,0601	1,1940	1,3148	1,4596
22	1,0784	1,2103	1,3290	1,4711
23	1,0958	1,2257	1,3425	1,4819
24	1,1122	1,2403	1,3551	1,4921
25	1,1277	1,2540	1,3671	1,5018
26	1,1425	1,2671	1,3784	1,5109
27	1,1566	1,2795	1,3891	1,5195
28	1,1700	1,2913	1,3994	1,5277
29	1,1828	1,3025	1,4091	1,5355
30	1,1950	1,3132	1,4183	1,5429
31	1,2067	1,3235	1,4272	1,5500
32	1,2180	1,3333	1,4356	1,5567
33	1,2287	1,3427	1,4437	1,5632
34	1,2390	1,3517	1,4515	1,5694
35	1,2490	1,3604	1,4589	1,5753
36	1,2585	1,3687	1,4661	1,5810
37	1,2677	1,3767	1,4730	1,5865
38	1,2766	1,3844	1,4796	1,5917
39	1,2851	1,3918	1,4860	1,5968
40	1,2934	1,3990	1,4921	1,6017
41	1,3014	1,4059	1,4980	1,6064
42	1,3091	1,4126	1,5038	1,6110
43	1,3166	1,4191	1,5093	1,6154
44	1,3238	1,4253	1,5147	1,6196
45	1,3308	1,4314	1,5198	1,6237

que indica cómo  $\hat{u}$  se distribuye normalmente <sup>4</sup> con parámetros  $(0, K \Omega K')$ . El estadístico  $d'$  [18] seguirá entonces una distribución teórica bajo la hipótesis nula del tipo.

$$P \{d' \leq d_\alpha \mid H_0\} = P \left\{ \sum_{i=1}^T \mu_i w_i^2 \leq 0 \right\} \quad [21]$$

donde  $\mu_i$  son ahora las raíces de la matriz  $(A - q_\alpha B_T) K K'$  <sup>5</sup> que sólo coincidirá con la de  $d$  si  $K K' = I$ , o lo que es igual si bajo la hipótesis nula resulta

$$\text{Var}(\hat{u}) = \sigma^2 K K' = \sigma^2 I \quad [22]$$

es decir, si la matriz de varianzas es escalar.

Cuando se selecciona como estimación de  $u$  alguna satisfaciendo tal condición de escalaridad en la matriz de varianzas, podrá aplicarse el contraste anteriormente citado, puesto que la región crítica de rechace será la misma y se obtendrá del cuadro 1. Si por el contrario la mencionada condición no se satisface, el contraste sólo podrá ser aplicado a partir de un nuevo cuadro que dependerá en cada caso de la matriz  $K$ , o en definitiva de la observación muestral que se maneja que es la que determina a ésta. El inconveniente en este caso, como es obvio, es el de tener que formar cada vez la correspondiente tabla, pero ello no es más que un inconveniente de tipo práctico y jamás una situación derivada de su inviabilidad teórica como muchas veces se ha apuntado.

La condición de escalaridad puede interpretarse como la exigencia para  $\hat{u}$  de la condición análoga a la que bajo la hipótesis nula satisface  $u$ , y es comprensible que así sea, puesto que de otro modo mal puede intentarse verificar la hipótesis de independencia en las distribuciones de  $u_i$  a partir de unas estimaciones que no lo son ni aun siéndolo aquéllas.

#### IV. ESTIMACIÓN DE LAS PERTURBACIONES

Admitido que a efectos de contrastar la hipótesis de ausencia de correlación, un camino cómodo es utilizar estimaciones  $u$  que satisfagan las condiciones:

$$\hat{u} = Ku \quad KK' = I$$

resulta conveniente averiguar si las habitualmente manejadas las satisfacen, y caso de que no sea así conocer al menos las posibilidades de aplicación que poseen al tema que nos ocupa.

4. La normalidad se sigue de [19 i] al ser  $u$  normal.

5. Véase el Corolario II del Apéndice.

El único modo conocido para estimar  $u$  consiste en hacerlo a partir de una estimación del modelo,  $\hat{\beta}$ , que supondremos averiguada, y tres serán las alternativas que consideraremos.

### A) Residuos ordinarios

Dado el modelo  $Y = X\beta + u$  una estimación «natural» de  $u$  es el vector de residuos  $e = Y - X\hat{\beta}$  que satisface, según puede verse en Johnston (1972, pp. 127-129 y 275-276), lo siguiente:

$$(i) \quad e = Mu \quad (ii) \quad \bar{e} = 0 \quad [23]$$

en donde

$$M = I_T - X(X'X)^{-1}X' \quad [24]$$

es una matriz de rango  $k$ , que cumple las propiedades de ser simétrica e idempotente, es decir:

$$MM' = M = M'$$

Al ser bajo la hipótesis  $H_0$ :

$$\text{Var}(e) = \sigma^2 MM' = \sigma^2 M$$

no se satisface la condición [22]. Procede, por tanto, seleccionar unas nuevas regiones críticas de rechazo, para cada nivel de significación, a partir de la distribución del estadístico [18] que en este caso, al ser  $B_T = I_T$  según [23 ii], tomará la forma

$$d_1 = \frac{e' Ae}{e' e} \quad [25]$$

cuya distribución bajo la hipótesis nula  $H_0$  depende de las raíces características no nulas  $\mu_t$  de la matriz  $(A - d_\alpha I_T) MM' = (A - d_\alpha I_T) M$ ,<sup>6</sup> con las cuales se obtienen como niveles de significación los siguientes

$$P\{d_1 \leq d_\alpha\} = P\left\{\sum_{t=1}^{T-k} \mu_t w_t^2 \leq 0\right\} \quad [26]$$

mediante los que se puede formar un contraste que consiste en aceptar la  $H_0$  si  $\hat{d}_1$  es menor que el valor  $d_\alpha$  deducido en la expresión [26] como corres-

6.  $(T - k)$  en total, puesto que éste es el rango de  $M$ .

pondiente a una determinada matriz  $M$ , y rechazarla en caso contrario. El contraste que acabamos de definir que coincide formalmente con el de Von Neumann lo notaremos en lo sucesivo con CRO (Contraste por Residuos Ordinarios).

Cuando se aúna la pesadez que supone la tabulación de [26] en cada caso concreto con el que los residuos  $e$  son las estimaciones de  $u$  más fácilmente alcanzables, y hasta época muy reciente las únicas conocidas, se comprende el porqué de los intentos realizados por multitud de autores con la idea de obtener un contraste alternativo basado en  $d_1$  que no tuviese en cuenta explícitamente a  $M$ . De entre todos ellos merecen destacarse los llevados a cabo por Durbin y Watson (1950-1951), los cuales propusieron un ingenioso procedimiento para contrastar  $H_0$ .

Los citados autores demuestran, en primer lugar, que la distribución de  $d_1$  está acotada superior e inferiormente por  $d_s$  y  $d_i$  con funciones de distribución respectivas

$$P\{d_s \leq d_{sa}\} = P\left\{\sum_{t=1}^{T-k} v_t^s w_t^2 \leq 0\right\}$$

$$P\{d_i \leq d_{ia}\} = P\left\{\sum_{t=1}^{T-k} v_t^i w_t^2 \leq 0\right\}$$

en las que  $v_t^s$  y  $v_t^i$  son raíces características de la ecuación  $(A - q_\alpha)$ .<sup>7</sup> Las primeras las  $T - k$  mayores, y las segundas las  $T - k$  menores excluidas las nulas. Dichas raíces no dependen evidentemente de  $X$ . A partir de lo anterior proceden a tabular los puntos más relevantes de la distribución de  $d_s$  y  $d_i$  aproximándolos mediante el ajuste a una distribución beta. Obtienen para cada par  $(T, k)$  y dado un nivel de significación  $\alpha$  el cuadro 2 que permite una utilización del siguiente tipo:

Si  $\hat{d}_1 < d_{ia}$ , lo cual implica que  $\hat{d}_1 \leq d_\alpha$ , se rechaza la hipótesis de ausencia de correlación, aceptándose la de que existe correlación positiva. Si  $\hat{d}_1 > d_{sa}$ , al ser también  $\hat{d}_1 > d_\alpha$ , se acepta la ausencia de ella. Si, por último,  $\hat{d}_1 \in (d_{ia}, d_{sa})$  nada puede asegurarse, pues también  $d_\alpha$  pertenece al mismo intervalo, con lo cual no queda claro el tipo de relación que liga a ésta con  $\hat{d}_1$ . En tal caso, la aplicación del contraste no da resultados determinados.

La hipótesis de ausencia de correlación negativa se contrasta de manera simétrica mediante  $4 - \hat{d}_1$ , dando lugar en conjunto a un procedimiento cuya

7. Von Neumann (1941) ha demostrado cómo tales raíces son de la forma

$$\lambda_t = 2 \left\{ 1 - \cos \frac{\pi(t-1)}{T} \right\} - q_\alpha$$



CUADRO 2

*Algunos limites de tolerancia del contraste Durbin Watson*

T	$\alpha = 1\% \quad k = 2$		$\alpha = 5\% \quad k = 3$	
	di	ds	di	ds
15	0,811	1,070	0,946	1,543
16	0,844	1,860	0,982	1,539
17	0,874	1,102	1,015	1,536
18	0,902	1,118	1,046	1,535
19	0,928	1,132	1,074	1,536
20	0,952	1,147	1,100	1,537
21	0,975	1,161	1,125	1,538
22	0,997	1,174	1,147	1,541
23	1,018	1,187	1,168	1,543
24	1,037	1,199	1,188	1,546
25	1,055	1,211	1,206	1,550
26	1,072	1,222	1,224	1,553
27	1,089	1,233	1,240	1,556
28	1,104	1,244	1,255	1,560
29	1,119	1,254	1,270	1,563
30	1,133	1,263	1,284	1,567
31	1,147	1,273	1,297	1,570
32	1,160	1,282	1,309	1,574
33	1,172	1,291	1,321	1,577
34	1,184	1,299	1,333	1,580
35	1,195	1,307	1,343	1,584
36	1,206	1,315	1,354	1,587
37	1,217	1,323	1,364	1,590
38	1,227	1,330	1,373	1,594
39	1,237	1,337	1,382	1,597
40	1,246	1,344	1,391	1,600
45	1,288	1,376	1,430	1,615
50	1,324	1,403	1,462	1,628
55	1,356	1,427	1,490	1,641
60	1,383	1,449	1,514	1,652
65	1,407	1,468	1,536	1,662
70	1,429	1,485	1,554	1,672
75	1,448	1,501	1,571	1,680
80	1,466	1,515	1,586	1,688
85	1,482	1,528	1,600	1,696
90	1,496	1,540	1,612	1,703
95	1,510	1,552	1,624	1,709
100	1,522	1,562	1,634	1,715

aplicación queda reflejada en el gráfico siguiente<sup>8</sup> en el que se representa la función de densidad de  $d_1$ :



Cuando el contraste anterior conocido con el nombre de sus autores, el cual notaremos en lo sucesivo con DW, es indeterminado, es aconsejable recurrir a lo que en principio intentaba evitarse, es decir, a obtener la distribución [26] de  $d_1$ , mediante el procedimiento de Imhof (1961) u otro alternativo propuesto por Pan Jai-Jian (1968) que recogen Durbin y Watson (1971) en un trabajo en el que intentan precisamente resolver esta cuestión. Otros métodos sugeridos por Hannan (1957), Theil y Nagar (1961) y Henshaw (1966), parecen desaconsejables, por una parte por faltarles en algunos casos generalidad, y por otra por estar basados en aproximaciones cuya precisión está lejos de ser óptima, por lo menos para tamaños muestrales reducidos.

A pesar de lo anterior, y para matizar adecuadamente lo dicho, es obligado indicar que Durbin (1970) ha propuesto un método que en principio parece más aceptable que los mencionados, para el caso de indeterminación. Su idea es la siguiente: Propone intercambiar el efecto que en el modelo producen las matrices  $X$  y  $L$ , donde esta última sería aquella para la cual la distribución del estadístico  $d_1$  calculada a través de [26] coincidiese con la del límite superior  $d_s$  del contraste DW. La matriz  $L$  resulta ser, como demostró Hannan (1960, pp. 128-130), la que está formada por los vectores característicos correspondientes a las  $k$  raíces menores de  $A$ , excluidas las nulas. La forma de intercambiar  $X$  por  $L$  se efectúa mediante una nueva regresión cuyos residuos a través de la Razón Modificada de Von Neumann, permiten aplicar un contraste cuya región crítica de rechace es  $d_1 \leq d_{sa}$ . Una exposición detallada de la nueva regresión se encuentra en Johnston (1972, pp. 253-254) y aquí no la consideramos con mayor detalle por entender que se aleja bastante de las líneas más recientes con las que se aborda este tema y que pasamos a considerar.

8. Tomado de Christ (1966, p. 529).

B) *Residuos óptimos con matriz escalar*

Ya hemos visto cómo  $d_1$  no es conveniente para contrastar la hipótesis  $H_0: \Omega = \sigma^2 I$ , por cuanto su distribución teórica depende de la observación de que se disponga, y ello no se resuelve totalmente por la aplicación del contraste DW al dar en muchos casos resultados indeterminados. Lo mejor es obtener estimaciones de  $u$  basadas en residuos que al contrario que  $e$  posean una matriz de varianzas escalar. Éste es el caso de los residuos óptimos con matriz escalar (BLUS en la terminología anglosajona, de Best Linear Unbiased Scalar Estimators) que notaremos con  $\tilde{u}$ . Han sido introducidos por Theil (1965) y Koerts (1967).

Johnston (1972, p. 255) demuestra cómo al cumplirse la condición de escalaridad sólo es posible estimar  $(T - k)$  componentes de  $u$ . Supongamos, sin pérdida de generalidad, que estimamos las  $(T - k)$  últimas (siempre es posible estimar otras sin más que reordenar adecuadamente los datos) y notemos con  $u_1$  el vector de perturbaciones correspondiente. Esta notación es conforme con la siguiente partición del modelo [4]:

$$\begin{bmatrix} Y_0 \\ Y_1 \end{bmatrix} = \begin{bmatrix} X_0 \\ X_1 \end{bmatrix} \beta + \begin{bmatrix} u_0 \\ u_1 \end{bmatrix} \quad [27]$$

donde los subíndices ceros hacen referencia a las  $k$  primeras observaciones y los unos a las  $(T - k)$  últimas contenidas en la muestra seleccionada. Si  $\tilde{u}_1$  simboliza el vector de estimaciones resultante, será:

$$\tilde{u}_1 = Cu \quad \text{con} \quad C = \|M_{11}^{-1/2} M_{10} : M_{11}^{1/2}\| \quad [28]$$

donde  $M_{10}$  y  $M_{11}$  son matrices deducidas de particionar  $M$  conforme a [27] del siguiente modo:

$$M = \begin{bmatrix} M_{00} : M_{01} \\ \vdots : \vdots \\ M_{10} : M_{11} \end{bmatrix} \quad [29]$$

donde, dada la idempotencia de  $M$ , resulta

$$(i) \quad M'_{10} = M_{01} \quad (ii) \quad M'_{11} = M_{11} \quad \text{y} \quad (iii) \quad M_{10}M_{01} + M_{11}M_{11} = M_{11}$$

Se pone en evidencia por lo anterior que  $u_1$  tiene una matriz de varianzas escalar bajo la hipótesis nula. En efecto, al ser según [28] nula la esperanza de  $u_1$  se cumplirá:

$$\text{Var}(u_1) = CC'$$

con

$$\begin{aligned}
 CC' &= \|M_{11}^{-1/2}M_{10} : M_{11}^{1/2}\| \quad \|M_{11}^{-1/2}M_{10} : M_{11}^{1/2}\|' = \\
 M_{11}^{-1/2}M_{10}M_{01}M_{11}^{-1/2} + M_{11} &= M_{11}^{-1/2}(M_{10}M_{01} + M_{11}M_{11})M_{11}^{-1/2} = \\
 M_{11}^{-1/2}M_{11}M_{11}^{-1/2} &= I_{T-k}
 \end{aligned}$$

Cuando se tiene presente que por su definición

$$MX = 0 \quad y \quad MY = e$$

de [29] resultarán:

$$\begin{aligned}
 M_{10}X_0 + M_{11}X_1 &= 0 \\
 M_{10}Y_0 + M_{11}Y_1 &= e_1
 \end{aligned} \tag{30}$$

de donde:

$$CX = M_{11}^{-1/2}M_{10}X_0 + M_{11}X_1 = M_{11}^{-1/2}(M_{10}X_0 + M_{11}X_1) = 0$$

y, además,

$$\begin{aligned}
 \tilde{u}_1 &= Cu = C(Y - X\beta) = CY = M_{11}^{-1/2}M_{10}Y_0 + M_{11}^{1/2}Y_1 = \\
 M_{11}^{-1/2}(M_{10}Y_0 + M_{11}Y_1) &= M_{11}^{-1/2}e_1
 \end{aligned} \tag{31}$$

que permite calcular los residuos lineales escalares óptimos en función de los ordinarios. Una vez obtenidos, el estadístico

$$d_2 = \frac{\tilde{u}'_1 A \tilde{u}_1}{\tilde{u}'_1 B_{T-k} \tilde{u}_1} \tag{32}$$

donde  $A$  y  $B_{T-k}$  son ahora matrices de orden  $(T - k)$ , sirve de base para aplicar un contraste que notaremos CRE, de Contraste por Residuos Escalares, cuya región crítica se obtendrá por el cuadro 1 en la que, como tamaño muestral, se toma  $T - k$  en lugar de  $T$ . Esto es así por cuanto según [21] la distribución de  $d_2$ , que es un caso particular de [18], sigue una distribución que al ser  $CC' = I_{T-k}$  depende de las  $T - k$  raíces de  $(A - q_\alpha B_{T-k})$  que son las utilizadas por Hart y Koerts y Abrahamse para construir la tabla a utilizar.

Una seria objeción que suele hacerse al contraste basado en los residuos escalares óptimos es la de que su aplicación implica un proceso muy largo. Y no andan desencaminados los que así opinan pues, según [31], se observa que el proceso a seguir es obtener los residuos ordinarios, de los que se seleccionan los  $T - k$  que corresponden a  $e_1$  y a partir de los cuales se forma la matriz  $M_{11}$

de las observaciones correspondientes. Luego se debe diagonalizar la inversa de tal matriz, a fin de obtener  $M_{11}^{-1/2}$ , y por último calcular el valor  $d_2$  del estadístico. Y todo ello al margen de que se da por sentado que cualquier conjunto de  $T - k$  observaciones, seleccionadas entre las  $T$ , es igualmente válido en posibilidades de utilización, cuando la verdad es que esto no es cierto ya que de [24] y [29] se sigue que:

$$M_{11}^{-1} = I + X_1 (X_0' X_0)^{-1} X_1'$$

con lo cual sólo podrán utilizarse aquellos conjuntos de observaciones entre los posibles, para los cuales  $X_0$  sea regular, ya que de otro modo el proceso quedaría invalidado al no existir la inversa de  $M_{11}$ . De entre todos los que satisfacen la condición anterior Theil (1971, pp. 217-218) propone seleccionar un conjunto de observaciones central en el que eliminando las  $k - m$  primeras y las  $m$  últimas resulte que la suma de las raíces de  $M_{11}$  sea máxima. Ello tiene su explicación en lo siguiente: Por una parte en que al ser central, con  $m$  pudiendo variar de 0 a  $k$ , se garantiza que las observaciones sean sucesivas, lo cual, a efectos de contrastar la correlación, parece básico, y por otra en que al cumplirse, según Koerts y Abrahamse (1971, p. 58), que:

$$E[(\tilde{u}_1 - u_1)'(\tilde{u}_1 - u_1)] = 2\sigma^2 \sum_{t=1}^{T-k} (1 - d_t) \quad [33]$$

donde  $d_t$  son las raíces de  $M_{11}$ , resulta que una manera de minimizar la suma esperada de errores de estimación medida por [33] es seleccionando la suma de raíces máxima. Por ello, una complicación adicional será la de escoger tal conjunto.

Abrahamse y Koerts (1971) y Dubbelman y otros (1972), establecen un método general de obtención de residuos, en el que los ordinarios y los escalares óptimos son casos particulares. Tales trabajos son relevantes en dos sentidos: en el apuntado, por lo que de síntesis supone, y en el de ofrecer la idea de obtener residuos ordinarios basados en matrices  $\bar{X}$  tipos, que son representativas de determinados fenómenos económicos, con lo cual la posibilidad de verificar la hipótesis  $H_0$  presupone solamente la existencia de algún criterio previo que permita clasificar la  $X$  correspondiente dentro del grupo representado por  $\bar{X}$ , del que se dispone previamente de la distribución exacta de su estadístico [25] y puede aplicarse CRO a partir de la tabla resultante.

### C) Residuos Recursivos

Otra forma alternativa de estimar  $u$  ha sido propuesta recientemente por Hedayat y Robson (1970), Phillips y Harvey (1973) y Harvey y Phillips (1974). Consiste en los llamados Residuos Recursivos («Recursive Residuals»

en la terminología original) los cuales, al igual que los escalares óptimos, sólo pueden obtenerse en número de  $T - k$  a partir de un conjunto de  $T$  observaciones. Notando con  $u_1^*$  al vector que los incluye, entendido como estimación de  $u_1$  conforme con la partición del modelo hecha en [27], resulta ser:

$$u_1^* = Ru \quad [34]$$

donde  $R$  es una matriz de orden  $(T - k) \times T$  definida de la siguiente manera:

$$R' = \|R_{k+1}, R_{k+2}, \dots, R_T\| \quad [35]$$

con

$$R'_h = \frac{1}{d_h} \left\| -X'_h (Z'_{h-1} Z_{h-1})^{-1} Z'_{h-1} : 1 : \Theta_{T-h} \right\| \quad [36]$$

en la que

$$d_h = \{1 + X'_h (Z'_{h-1} Z_{h-1})^{-1} X_h\}^{1/2} \quad b = k + 1, \dots, T \quad [37]$$

siendo  $X'_h$  la fila  $h$ -ésima de la matriz de observaciones  $X$ ,  $Z_{h-1}$  la sub-matriz formada por las  $(h - 1)$  primeras filas de  $X$  que satisface, por tanto:

$$(i) \quad Z'_h = \|Z'_{h-1} : X_h\| \quad (ii) \quad Z_T = X$$

y  $\Theta_{T-h}$  un vector fila de  $T - h$  ceros.

Es importante probar que la matriz de las varianzas de  $u_1^*$  es escalar, es decir,

$$\text{Var}(u_1^*) = \sigma^2 R R' = \sigma^2 I \quad [38]$$

lo cual equivale a probar que  $R R' = I$ . En efecto:

$$R R' = \left\| \begin{array}{cccc} R'_{k+1} R_{k+1} & R'_{k+1} R_{k+2} & \dots & R'_{k+1} R_T \\ R'_{k+2} R_{k+1} & R'_{k+2} R_{k+2} & \dots & R'_{k+2} R_T \\ \dots & \dots & \dots & \dots \\ R'_T R_{k+1} & R'_T R_{k+2} & \dots & R'_T R_T \end{array} \right\|$$

con

$$R'_h R_h = \frac{1}{d_h^2} [X'_h (Z'_{h-1} Z_{h-1})^{-1} Z'_{h-1} Z_{h-1} (Z'_{h-1} Z_{h-1})^{-1} X_h + 1] =$$

$$\frac{1}{d_h^2} [X'_h (Z'_{h-1} Z_{h-1})^{-1} X_h + 1] = 1$$

y, si  $b < n$ ,

$$R'_n R_n = \frac{1}{d_n d_n} [X'_h (Z'_{h-1} Z_{h-1})^{-1} Z'_{h-1} Z_{h-1} (Z'_{n-1} Z_{n-1})^{-1} X_n - \\ X'_h (Z'_{n-1} Z_{n-1})^{-1} X_n] = 0$$

con lo que también, si  $n > b$ , resulta  $R'_n R_h = 0$ . De donde se cumple  $RR' = I$  como queríamos probar. Como sea que, además,

$$R'_h X = \left\| \begin{array}{c} -X'_h (Z'_{h-1} Z_{h-1})^{-1} Z'_{h-1} : 1 : \Theta_{T-H} \end{array} \right\| \left\| \begin{array}{c} Z_{h-1} \\ \dots \\ X'_h \\ \dots \\ Z_h^c \end{array} \right\| = 0$$

$$b = k + 1, k + 2, \dots, T$$

es decir,  $RX = 0$ , los residuos recursivos pueden obtenerse mediante la relación

$$u_1^* = Ru = R(Y - X\beta) = RY \quad [39]$$

Si notamos con  $W_h$  la submatriz de  $Y$  formada por sus  $b$  primeros elementos, es decir,

$$W'_h = \left\| W'_{h-1} : Y_h \right\| \quad \text{con } W_T = Y$$

se cumplirá, de acuerdo con [39],

$$u_i^* = R'_i Y = \frac{1}{d_h} [-X'_h (Z'_{h-1} Z_{h-1})^{-1} Z'_{h-1} W_{h-1} + Y_h] \\ - \frac{1}{d_h} [Y_h - X'_h \hat{\beta}_{h-1}] \quad b = k + 1, \dots, T \quad [40]$$

en donde  $\hat{\beta}_{h-1}$  es la estimación de  $\beta$  obtenida a partir de las  $b - 1$  primeras observaciones incluidas en la muestra. Resulta, por tanto, que el cálculo de los residuos queda reducido a determinar los sucesivos valores de  $\hat{\beta}_h$  y sustituirlos adecuadamente en [40]. Por fortuna, según demuestran los autores citados y de ahí el nombre de los residuos, el cálculo puede hacerse por recurrencia a través de la relación

$$\hat{\beta}_h = \hat{\beta}_{h-1} + \frac{(Z'_{h-1} Z_{h-1})^{-1} X_h (Y_h - X'_h \hat{\beta}_{h-1})}{d_h^2} \quad [41]$$

que permite un tratamiento muy cómodo mediante la utilización de un ordenador, quedando reducida la aplicación de [41] al conocimiento de

$$\hat{\beta}_k = (Z_k)^{-1} W_k$$

que es la estimación del modelo  $W_k = Z_k\beta + u_k$  formado por  $k$  ecuaciones. Bien entendido que además se satisface también

$$(Z'_k Z_k)^{-1} = (Z'_{h-k} Z_{h-k})^{-1} - \frac{(Z'_{h-1} Z_{h-1})^{-1} X_h X'_h (Z'_{h-1} Z_{h-1})^{-1}}{d_h^-} \quad [42]$$

Vemos, pues, que los residuos recursivos poseen similares propiedades que los escalares óptimos. Son, según [34] y [38], lineales, insesgados y se distribuyen normalmente con parámetros  $(0, R \Omega R')$ . La única diferencia es que no son óptimos en el sentido de que, según prueba Theil (1971, pp. 209-213), la expresión [33] es mínima y traduce, con independencia de las  $d_t$  que se seleccionen, que las  $\tilde{u}_1$  son óptimas entre las de su clase. Siendo así, claro está que los residuos recursivos no podrán serlo, pues no coinciden con los escalares óptimos al ser  $R \neq C$  como fácilmente se comprueba al compararlos. A pesar de ello, interesa resaltar que sí se satisfacen las propiedades que, bajo la hipótesis nula, implican para el estadístico [18], que en este caso toma la forma particular:

$$d_3 = \frac{u_1^{*'} A u_1^*}{u_1^{*'} B_{T-k} u_1^*} \quad [43]$$

el que pueda servir de base para la aplicación de un contraste cuya región crítica de rechace para  $H_0$  es la derivada del cuadro 1, en la que nuevamente como tamaño muestral debe tomarse  $T - k$ . Dicho contraste lo notaremos en lo sucesivo con CRR, de Contraste por Residuos Recursivos.

## V. COMPARACIÓN ENTRE LOS CONTRASTES

En el apartado anterior se han definido cuatro contrastes distintos para la hipótesis de ausencia de correlación en el término de perturbación de un modelo econométrico. Para cada uno de ellos se ha considerado su menor o mayor facilidad de aplicación. En este sentido parece innegable que el considerado en último lugar, el CRR, resulta el más interesante, puesto que es el más sencillo de manejar, al margen del DW que tiene el gran inconveniente de presentar regiones indeterminadas. Ahora bien, el decantarse sin más por él parece una conclusión precipitada, toda vez que lo que de verdad define la calidad de un contraste, como pone de relieve Arnaiz (1965, pp. 576 y 604-



609), es su función de potencia. Si, como cabe esperar, lo que realmente se pretende es contrastar válidamente la hipótesis, sólo cuando tal función sea sensiblemente igual para dos o más contrastes podrán ser otras consideraciones, entre ellas las de simplicidad, las que decanten la elección.

La función de potencia de un contraste tiene que ver con su capacidad para rechazar la hipótesis nula que se considere, en este caso  $\Omega = \sigma^2 I$ , y viene definida en términos de probabilidad. La potencia para un determinado punto hay que entenderla como el valor que toma la función anterior en él. Dicho de otro modo, la potencia de un contraste para  $\Omega = \Omega_1$  es la probabilidad de que utilizándolo se rechace la hipótesis nula. Es perfectamente comprensible la importancia de tal noción ya que en base a ella un contraste será tanto más válido cuanto menor sea su potencia para  $H_0$  y mayor lo sea para los puntos que componen la hipótesis alternativa  $H_1$ .

Partiendo de que por la forma como se seleccionan las regiones críticas para los diversos estimadores que hemos considerado está garantizada la misma potencia  $\alpha$  —equivalente al nivel de significación— bajo la hipótesis nula, resultará que la comparación deberá basarse únicamente en la potencia que cada uno muestre para los posibles valores de  $H_1$ . Ahora bien, como la potencia relativa está en función de la especificación concreta de  $\Omega$  que se considere, una cuestión previa será pronunciarse sobre la estructura que ésta posee bajo la mencionada hipótesis. A tal fin recordemos que en el apartado II se puso de manifiesto cómo, si las perturbaciones están correlacionadas, el esquema habitualmente aceptado para expresarlas es el «autorregresivo de primer orden», correspondiente a un proceso de generación del tipo [8]. Pues bien, según demuestra Alcaide (1966, pp. 196-197), dicho esquema determina para el vector  $u$  de perturbaciones, una distribución con parámetros:

$$E(u) = 0 \quad \text{Var}(u) = \Omega = \begin{vmatrix} 1 & \rho & \rho^2 & \vdots & \rho^{T-1} \\ \rho & 1 & \rho & \vdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \vdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \vdots & 1 \end{vmatrix} \quad [44]$$

cuya normalidad está garantizada de ser normales las variables  $\varepsilon_t$  que determinan linealmente el comportamiento de las  $u_t$ .

Como consecuencia de lo anterior, pronunciarse sobre el punto  $\Omega = \Omega_1$  será equivalente a hacerlo sobre  $\rho = \rho_1$ , y esto es precisamente lo que haremos en lo que sigue, sobre la base de que, excepción hecha del contraste DW (al que después nos referiremos), la potencia de los restantes vendrá determinada por la distribución del estadístico de Von Neumann modificado que puede expresarse como sigue:

$$P\{d' \leq d_\alpha / \rho = \rho_1\} \quad [45]$$

en la que  $d'$  denota genéricamente tal estadístico y  $d_\alpha$  el límite superior de la región crítica asociada a  $\alpha$ , que no es igual para todos. Para CRE y CRR es el señalado en el cuadro 1 y para CRO es el que aparecería en [26] para una determinada matriz de observaciones  $X$ .

El proceso de obtención de la potencia culmina en este caso al comprobar cómo, según el Corolario II del Apéndice, la probabilidad [45] podrá calcularse mediante la expresión [A.2], a partir de las raíces de las matrices.

$$\begin{array}{ll} (A - d_\alpha I_T) M \Omega M & \text{para } d_1 \text{ de CRO} \\ (A - d_\alpha B_{T-k}) C \Omega C' & \text{para } d_2 \text{ de CRE} \\ (A - d_\alpha B_{T-k}) R \Omega R' & \text{para } d_3 \text{ de CRR} \end{array} \quad [46]$$

que, como se comprueba, dependen en cualquier caso —aparte del valor seleccionado para  $\rho_1$  y del nivel de significación  $\alpha$ — de las matrices  $X$  de valores observados. Todo lo anterior en virtud de que las matrices de varianzas de  $e$ ,  $\tilde{u}_1$  y  $u_1^*$  bajo la hipótesis  $H_1$  son, respectivamente,  $M \Omega M$ ,  $C \Omega C'$  y  $R \Omega R'$  con  $\Omega$  tomada de [44].

Por su parte, para obtener la potencia del contraste DW bajo la hipótesis  $H_1$ :  $\rho = \rho_1$  deberá calcularse la probabilidad [45] en la que  $d_\alpha$  sea el límite inferior de la región de indeterminación, es decir, el  $d_{ix}$  que aparece en el cuadro 2 para un nivel de significación  $\alpha$ , y  $d'$  el estadístico  $d_1$ . También por la misma razón que antes la distribución de [45] dependerá de los valores observados, a través de las raíces características de la matriz

$$(A - d_{ix} I_T) M \Omega M \quad [47]$$

Existe, sin embargo, una matización a realizar en este caso y es la de que la potencia calculada estará claramente mediatizada por la amplitud de la región ( $d_{ix}$ ,  $d_{sx}$ ). Por ello es imprescindible complementar cualquier análisis de la potencia de DW con una medida del grado bajo el cual el contraste se muestra indeterminado, lo cual puede conseguirse averiguando la probabilidad

$$\begin{aligned} P \{d_{ix} < d_1 \leq d_{sx} / \rho = \rho_1\} &= P \{d_1 \leq d_{sx} / \rho = \rho_1\} - \\ &- P \{d_1 \leq d_{ix} / \rho = \rho_1\} \end{aligned} \quad [48]$$

en la que el sustraendo es la potencia y el minuendo la probabilidad de no cometer un error del tipo II, es decir, de no aceptar  $H_0$  cuando es falsa.<sup>9</sup> Tal probabilidad se obtiene aplicando [45] a partir de las raíces de la matriz

$$(A - d_{sx} I_T) M \Omega M \quad [49]$$

9. Nótese que para cualquier otro contraste la diferencia sería nula, pues, en general, la probabilidad de no cometer un error del tipo II y la potencia coinciden.

Antes de proseguir es conveniente indicar que para calcular la potencia de los contrastes bajo la hipótesis  $H_1: \rho = \rho_1 < 0$  debe procederse a determinar las probabilidades

$$P\{d' \geq 4 - d_\alpha / \rho = \rho_1\} = 1 - P\{d' \leq 4 - d_\alpha / \rho = \rho_1\} \quad [50]$$

en las que  $d'$  y  $d_\alpha$  tienen la misma interpretación que en [45]. La probabilidad que aparece como sustraendo en el segundo miembro puede hallarse aplicando un procedimiento análogo al seguido en el caso de correlación positiva cuidando de sustituir en [46]  $d_\alpha$  por  $4 - d_\alpha$  y en [47] y [48]  $d_{ia}$  y  $d_{sa}$  por  $4 - d_{ia}$  y  $4 - d_{sa}$ , respectivamente. No es cierto, por otra parte y a pesar de que frecuentemente se encuentra citado en la literatura sobre el tema, que para la correlación negativa sea también de aplicación la expresión [45] por cuanto la simetría que lo justificaría sólo se da bajo la hipótesis nula.

Hay que resaltar que por todo lo dicho hasta aquí la potencia comparativa de los contrastes que manejamos dependerá de una serie de causas que van desde la matriz  $X$  de observaciones que se considere hasta el nivel  $\alpha$  seleccionado, pasando por el coeficiente  $\rho_1$  que se tome, el número de variables del modelo e implícitamente a través de  $d_\alpha$  del tamaño muestral  $T$ .<sup>10</sup> De ahí que sean varios los trabajos publicados tratando de aclarar la situación, pues no es fácil ponerse de acuerdo sobre la incidencia conjunta de todas ellas. En lo que sigue intentaremos hacer un resumen de la situación en que se halla actualmente la investigación sobre esta materia. Los trabajos que se citen obtienen sus conclusiones a partir de las potencias calculadas mediante las expresiones [45] y [50] siguiendo el procedimiento de computación de Imhof (1961).

A partir de unas observaciones sobre variables económicas tomadas de la Oficina Central de Planificación Holandesa, Abrahamse y Koerts (1969) establecen la potencia comparativa entre los contrastes CRO, DW y CRE para  $\alpha = 0,05$  y  $\rho = 0,8$  que se muestra en el cuadro 3. La importancia de él está en que ninguna de las experiencias realizadas posteriormente por éstos u otros autores está en contradicción con las conclusiones que de él pueden derivarse y que son las siguientes: En primer lugar se observa que la potencia de CRO es sensiblemente mayor que la de los otros dos. Esto no puede sorprender al tener en cuenta su menor restrictividad, pues está basado en estimaciones de  $u$  que son cuantitativa y cualitativamente mejores que las que sirven de base a CRE, y por otra parte no posee ninguna región de indeterminación. En segundo lugar está que la potencia de CRE es mayor que la de DW, pero sólo hasta un determinado tamaño muestral ( $T = 25$ ) en el que ya coinciden, y en donde, además, ambas potencias se acercan a las de CRO. Marginalmente puede verse también cómo el número de variables incluidas en el modelo influye negativa-

10. La potencia de CRR depende también de la selección que se haga de las  $k$  perturbaciones que no se estiman, ya que no hay nada escrito por ahora sobre cuál es la mejor.

CUADRO 3

 $\alpha = 0,05$  $\rho = 0,8$ 

Tamaño muestral	Número de variables (1) explicativas	Potencias			Prob. de indetermin. en DW
		CRO	CRE	DW	
15	2	0,80	0,73	0,74	0,13
	3	0,73	0,63	0,52	0,37
	4	0,69	0,55	0,34	0,58
	5	0,67	0,45	0,23	0,71
21	2	0,93	0,90	0,90	0,05
	3	0,91	0,86	0,84	0,12
	4	0,90	0,82	0,77	0,20
	5	0,89	0,77	0,67	0,31
25	2	0,97	0,95	0,95	0,03
	3	0,96	0,93	0,93	0,05
	4	0,95	0,89	0,89	0,10
	5	0,95	0,83	0,83	0,06

(1) Se incluye como variable explicativa el término independiente.

mente en todas las potencias, cuestión que parece explicable por la incidencia de la multicolinealidad en las estimaciones de  $u$ .

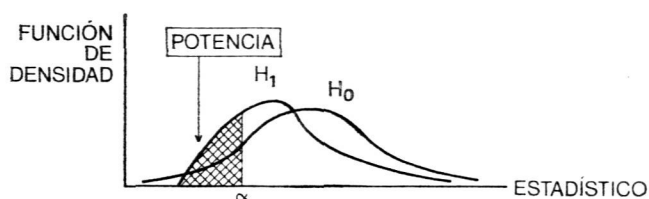
A la vista de lo anterior una estrategia recomendada por los autores del cuadro consiste en que si la diferencia  $T - k$  es elevada (mayor de 25) lo más cómodo es aplicar DW, pues los resultados serán sensiblemente iguales a los deducidos de utilizar CRE, que al ser más complicado puede desecharse,<sup>11</sup> y aun cuando serán menos significativos que los deducidos de CRO, la diferencia no justifica lo engorroso que es aplicar este último.

Cuando  $T - k$  sea menor de 25, lo cual tratándose —como comúnmente ocurre— de modelos temporales es por desgracia más frecuente de lo que sería de desear, el camino válido para contrastar la ausencia de correlación en el término de perturbación es utilizar CRO. Sólo cuando la magnitud del trabajo que esto implica haga desistir al investigador en su aplicación procederá la utilización de otro contraste, el cual, sin duda, deberá ser CRE y nunca DW.

Una conclusión que también parece posible deducir del cuadro 3 es que las potencias de todos los contrastes son relativamente elevadas. Hay que indicar desde este momento que posiblemente ello sea debido a los valores de  $\alpha$  y  $\rho$  manejados, que son singularmente elevados. Para justificar esta afirmación

11. Salvo en el caso de indeterminación.

consideremos el siguiente gráfico en el que se representa el modo como la correlación positiva afecta a la distribución del estadístico de Von Neumann modificado.



La asimetría que en él se observa está motivada por el valor que tome  $\rho$ , al afectar éste al sustraendo que aparece en el numerador del estadístico. Cuanto mayor sea, menor será tal numerador y mayor en consecuencia la asimetría. Al ser así, valores elevados de  $\rho$  determinarán, fijo  $\alpha$ , mayores potencias que valores bajos. Como por otra parte, dada una distribución correspondiente a un valor de  $\rho$ , el que sea  $\alpha$  mayor o menor también influye al variar la zona rayada en el gráfico, parece confirmarse la explicación de los altos valores obtenidos.<sup>12</sup> En lo que antecede no se ha mencionado la influencia que la matriz  $X$  tiene sobre la potencia, cuando está claro por [46] que ésta existe. Una posible forma de explicar su efecto es a través del mayor o menor apuntamiento en la distribución, como consecuencia de la forma concreta de  $X$ , fijados los valores de  $\alpha$  y  $\rho$ . En el experimento descrito hay razones sobradas para pensar que ésta es otra causa más de las elevadas potencias obtenidas, sobre todo cuando se las compara con las que más adelante veremos se obtienen para matrices distintas en igualdad de valores para las restantes variables.

Koerts y Abrahamse (1971, pp. 96-97) y Phillips y Harvey (1973) obtienen unos resultados que resumidos dan lugar al cuadro 4, a partir de una matriz de

CUADRO 4  
 $\alpha = 0,05$        $T = 15$        $k = 3$

$\rho$	CRE	CRR	DW	IND.
0,2	0,10	0,11	0,01	0,14
0,5	0,23	0,26	0,07	0,31
0,8	0,49 (°)	0,45	0,29 (°)	0,36
-0,2	0,12	0,12	—	—
-0,5	0,34	0,35	—	—
-0,8	0,64	0,62	—	—

observaciones cuyos elementos están generados por un mecanismo que elimina cualquier vestigio de dependencia entre los valores de las diversas variables seleccionadas.

12. En principio no parecen existir razones para pensar que lo dicho no deba ser aceptado para cualesquiera que sean el contraste y los valores de  $T$  y  $k$  seleccionados.

De él se pueden extraer también varias conclusiones. La primera y más importante que está en la línea de lo ya apuntado. Incluso en cuanto a que la potencia es creciente con  $\rho$  y que la matriz  $X$  tiene su influencia, como se comprueba al comparar las señaladas ( $^{\circ}$ ), con sus correspondientes del cuadro 3. Pero lo más importante es que en él aparecen comparados, a través de su potencia, CRE y CRR, y los resultados en este sentido no pueden ser menos orientativos a la hora de pronunciarse por alguno de ellos.<sup>13</sup> De ahí que pueda afirmarse que en principio resulta indiferente su utilización.<sup>14</sup> Cuando esto se relaciona con el distinto grado de dificultad en su determinación respectiva, la elección entre ambos es cuando menos dudosa, dependiendo en último extremo del grado de meticulosidad de quien deba utilizarlos.

## VI. CONCLUSIONES

En el apartado II se puso de manifiesto que la utilización de contrastes de ausencia de correlación serial en el término de perturbación perseguía como finalidad inmediata el disponer de un indicador previo que de ser aceptable evitase la utilización de otros métodos de estimación que allí se mencionaban como alternativos al uso de MCO. Estamos ahora en condiciones de asegurar que un planteamiento aceptable es el de aplicar como tales, DW si el conjunto de observaciones es grande, y CRR si es pequeño o si aquél se muestra indeterminado, ya que enzarzarse en manejar cualquier otro de los descritos no mejora excesivamente las conclusiones que de los mencionados se derivan, salvo CRO para tamaños pequeños. Pero en este caso no queda demasiado claro que la complicación que viene asociada a su manejo no sea superior a la que se asocia a los métodos de estimación que se intentaban evitar.

Cuando, aplicado un contraste, la ausencia de correlación no pueda ser rechazada al nivel de significación seleccionado serán suficientemente válidas las estimaciones previamente obtenidas por MCO, mayormente cuando se tiene presente que una aceptación indebida es más frecuente para valores de  $\rho$  pequeños, bajo los cuales ya se dijo que los métodos alternativos no mejoraban prácticamente nada las estimaciones por MCO.

Según lo visto en las páginas anteriores resulta que incluso para valores de  $\rho$  realmente elevados la potencia de los contrastes no es demasiado grande cuando el número de observaciones es escaso. Esto que es un inconveniente evidente, pues se traduce en un elevado índice de aceptación de  $H_0$  cuando

13. Los resultados se han obtenido eliminando las tres primeras observaciones en cada caso, con lo cual si bien al procederse de este modo no está garantizado que aparezcan las mejores potencias para CRE, no es menos cierto que tampoco está probado que sí lo sean las de CRR, sino más bien lo contrario, pues a pesar de que en este sentido Phillips y Harvey demuestran que no hay regla fija de selección, aconsejan tomar las centrales.

14. La verdad es que en estudios más completos a partir de diferentes  $X$  los autores anteriores obtienen que, aproximadamente, la eficiencia de CRR es 0,9 en relación con CRE.

es falsa, no parece tener solución en la fase en que actualmente se encuentra la investigación sobre el tema. Sólo la especificación de nuevos contrastes, por ahora desconocidos, que fuesen más potentes podría resolverlo.

Es necesario realizar una última precisión. Todos los contrastes están basados en una hipótesis alternativa  $H_1$  en forma de un proceso autorregresivo de primer orden. Cuando esto no es cierto la potencia de los contrastes se ve sensiblemente disminuida, lo mismo que cuando entre las variables explicativas del modelo [1] hay alguna retardada. Ambas situaciones que han sido últimamente consideradas por diversos autores a un nivel que queda reflejado en Sims (1974, pp. 319-324) aquí no se considerarán por entender que ello es materia suficiente para dar contenido a otro trabajo similar al presente.

## VII. APÉNDICE \*

El objetivo del presente apéndice es desarrollar una serie de cuestiones cuya utilización ha sido continua en las páginas anteriores y que a pesar de ello hemos creído preferible desligar del contenido fundamental por entender que procediendo de este modo se facilitaba la comprensión separada de aquél. El Teorema que sirve de base ha sido tomado, con la notación adaptada al presente caso y con las matizaciones oportunas, de Koerts y Abrahamse (1971, pp. 81-89).

\* Una vez terminado de escribir este trabajo acaba de aparecer publicado un artículo (L'ESPERANCE, W. L., y TAYLOR, D., «The Power of Four Tests of Autocorrelation in the Linear Regression Model», *Journal of Econometrics*, 3 (1975), 1-21) cuyas concomitancias con el presente hacen necesario le dediquemos esta nota. Se trata en él de comparar la potencia de cuatro contrastes distintos bajo diferentes hipótesis de comportamiento. Para ello sus autores se basan en la notación y terminología derivadas de la sistematización efectuada por Abrahamse y Koerts (1971). Los cuatro contrastes que se comparan son según su notación OLS, BLUS, AL y D. Los dos primeros son precisamente los que hemos venido llamando CRO y CRE, y en cuanto a los otros dos conviene hacer una serie de precisiones. El AL es un contraste (propuesto por ABRAHAMSE, A. P. J., y LOUTER, A. S., «On a New Test for Autocorrelation in Least Squares Regression», *Biometrika*, 58 (1971), 53-60) que si no ha sido tomado en consideración en nuestro trabajo es por la razón muy evidente de estar basado en ciertas hipótesis restrictivas sobre el comportamiento de las variables del modelo, del tipo de las que sirvieron de base en su día a los contrastes de Theil y Nagar (1961) y Hannan (1957). En realidad es una versión sistematizada de ellos y en consecuencia no cabía esperar para su potencia resultados mejores que los que L'Esperance y Taylor han obtenido, los cuales les excluyen de toda posible utilización.

Otra cuestión distinta es el contraste D, que en realidad es el basado en la transformación de los residuos propuesta por Durbin (1970). Ya apuntábamos en la página 212 que sus posibilidades parecían más elevadas. Si a pesar de ello no se manejó a la hora de establecer las comparaciones fue única y exclusivamente por entender que estaba relegado a un segundo plano. Nadie se había atrevido a recoger el reto que su autor lanzaba desde las páginas del artículo que lo definía, en el sentido de comparar su potencia con la de CRE. Parece que esto es lo que ahora han hecho los autores que consideramos en esta nota y lo realmente importante es que para valores de  $\rho$  positivos su potencia es mayor que la de CRE. Parece, pues, que en el futuro habrá que tomarlo en consideración.

Una última cuestión que se destaca en el trabajo comentado es que salvo lo anterior está en total concordancia con el esquema propuesto en el presente.

TEOREMA. — Sea  $\xi$  una variable aleatoria  $n$ -dimensional con distribución normal de parámetros  $(0, \Delta)$  y sean  $\xi'A\xi$  y  $\xi'B\xi$  dos formas cuadráticas de matrices  $A$  y  $B$ , la segunda definida positiva. La distribución de la variable

$$\eta = \frac{\xi'A\xi}{\xi'B\xi} \quad [\text{A.1}]$$

tiene como Función de Distribución la siguiente:

$$F_\eta(z) = P(\eta \leq z) = P\{\sum_1^n \lambda_i \gamma_i^2 < 0\} \quad [\text{A.2}]$$

donde  $\gamma_i^2$  son variables aleatorias con distribución « $j$ -cuadrado» con un grado de libertad, y  $\lambda_i$  las raíces características de la matriz

$$(A - zB) \Delta \quad [\text{A.3}]$$

En efecto:

$$\begin{aligned} F_\eta(z) = P(\eta \leq z) &= P\left(\frac{\xi'A\xi}{\xi'B\xi} \leq z\right) = P(\xi'A\xi \leq \xi'zB\xi) = \\ &= P(\xi'(A - zB)\xi \leq 0) \end{aligned}$$

en donde definiendo una nueva variable  $\psi = \Delta^{-1/2} \xi$ ,<sup>15</sup> que será normal de parámetros:

$$\begin{aligned} E(\psi) &= E(\Delta^{-1/2} \xi) = \Delta^{-1/2} E(\xi) = 0 \quad \text{y} \\ \text{Var}(\psi) &= E(\Delta^{-1/2} \xi \xi' \Delta^{-1/2}) = \Delta^{-1/2} E(\xi \xi') \Delta^{-1/2} = \Delta^{-1/2} \Delta \Delta^{-1/2} = I \end{aligned}$$

resultará:

$$F_\eta(z) = P(\psi' \Delta^{1/2} (A - zB) \Delta^{1/2} \psi \leq 0) = P(\psi' Q' \Lambda Q \psi \leq 0)$$

donde  $\Lambda = Q \Delta^{1/2} (A - zB) \Delta^{1/2} Q'$  es la matriz diagonalizada de  $\Delta^{1/2} (A - zB) \Delta^{1/2}$  y  $Q$  es la de vectores característicos, con  $QQ' = I$ . Si definimos ahora una nueva variable aleatoria  $\gamma = Q\psi$ , normal, de parámetros:

$$\begin{aligned} E(\gamma) &= E(Q\psi) = QE(\psi) = 0 \quad \text{y} \\ \text{Var}(\gamma) &= E(Q\psi\psi'Q') = QE(\psi\psi') = QQ' = I \end{aligned}$$

se cumple, como queríamos probar, que

$$F_\eta(z) = P(\gamma' \Lambda \gamma \leq 0) = P(\sum_1^n \lambda_i \gamma_i^2 \leq 0)$$

15. La existencia de  $\Delta^{-1/2}$  está garantizada por ser  $\Delta$  simétrica y diagonalizable.



donde  $\lambda_i$  son los elementos de  $\Lambda$ , es decir, las raíces características de

$$\Delta^{1/2} (A - zB) \Delta^{1/2}$$

que coinciden con las de  $(A - zB) \Delta$  en virtud de que tales raíces son invariantes respecto a la conmutatividad conforme en el producto de matrices.  $\gamma_i$  son los elementos del vector aleatorio  $\gamma$  por lo que cumplirán ser variables normales reducidas, y sus cuadrados «jis-cuadrado» con un grado de libertad. COROLARIO I. — Si  $\xi$  es una variable aleatoria  $n$ -dimensional con parámetros  $(0, \sigma^2 I)$ , la distribución de probabilidad de (A.1) tendrá como Función de Distribución:

$$F_n(z) = P(\eta \leq z) = P(\sum_1^n \mu_i \gamma_i^2 \leq 0) \quad [\text{A.4}]$$

donde  $\gamma_i^2$  tienen la misma significación que antes, y  $\mu_i$  son en este caso las raíces de la matriz:

$$(A - zB) \quad [\text{A.5}]$$

En efecto, según [A.2], la Función debería ser

$$F_n(z) = P(\sum_1^n \lambda_i \gamma_i^2 \leq 0)$$

donde  $\lambda_i$  son las raíces de  $(A - zB) \sigma^2 I = (A - zB) \sigma^2$ , pero éstas son de la forma  $\sigma^2 \mu_i$ , donde  $\mu_i$  son las de la matriz [A.5], resultando

$$\begin{aligned} F_n(z) &= P(\sum_1^n \lambda_i \gamma_i^2 \leq 0) = P(\sum_1^n \sigma^2 \mu_i \gamma_i^2 \leq 0) = \\ &= P(\sigma^2 \sum_1^n \mu_i \gamma_i^2 \leq 0) = P(\sum_1^n \mu_i \gamma_i^2 \leq 0) \end{aligned}$$

como queríamos demostrar.

COROLARIO II. — Si  $\xi$  es una variable aleatoria  $n$ -dimensional de parámetros  $(0, C\Omega C')$  la Función de Distribución de [A.1] bajo estos supuestos será [A.2], donde  $\lambda_i$  son las raíces de la matriz  $(A - zB) C\Omega C'$ , las cuales, bajo la hipótesis  $\Omega = \sigma^2 I$ , se transforman en las de

$$(A - zB) CC' \quad [\text{A.6}]$$

y si además se cumple  $CC' = I$ , vuelven a ser las de [A.5].

La demostración de ello es inmediata a partir del Corolario I.

## VIII. BIBLIOGRAFÍA

1. ABRAHAMSE, A. P. J., «A Comparison between the Power of the Durbin-Watson Test and the Power of the BLUS Test», *J.A.S.A.*, 64, 1969, 938-948.
2. ABRAHAMSE, A. P. J., y KOERTS, J., «New Estimators of Disturbances in Regression Analysis», *J.A.S.A.*, 66, 1971, 71-74.
3. AITKEN, A. C., «On Least Squares and Linear Combination of Observations», *Proc. Royal Society of Edinburgh*, 55, 1934, 42-48.
4. ALCAIDE, A., *Lecciones de Econometría, Métodos Estadísticos*, Copigraf, Madrid, 1966.
5. AMES, E., y REITER, S., «Distributions of Correlation Coefficients in Economic Time Series», *J.A.S.A.*, 56, 1961, 637-656.
6. ANDERSON, T. W., «Least Squares and Best Unbiased Estimates», *Annals of Mathematical Statistics*, 33, 1962, 266-272.
7. ARNAIZ, G., *Introducción a la Estadística Teórica. I*, Lex Nova, Valladolid, 1965.
8. COCHRANE, D., y ORCUTT, G. H., «Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms», *J.A.S.A.*, 44, 1949, 32-61.
9. CHRIST, C. F., *Econometric Models and Methods*, John Wiley and Sons, Inc., Nueva York, 1966. (Existe una versión castellana de Ed. Limusa que es ininteligible.)
10. DUBBELMAN, D.; ABRAHAMSE, A. P. J., y KOERTS, J., «A New Class of Disturbance Estimators in the General Linear Model», *Statistica Neerlandica*, 4, 1972, 127-141.
11. DURBIN, J., «An alternative to the Bounds Test for Testing for serial Correlation in Least Squares Regression», *Econometrica*, 38, 1970, 422-429.
12. DURBIN, J., «Estimation of Parameters in Time Series Regression Models», *Journal of the Royal Statistical Society, series B*, 20, 1960, 139-153.
13. DURBIN, J., y WATSON, G. S., «Testing for Serial Correlation in Least Squares Regression, I», *Biometrika*, 37, 1950, 409-428.
14. DURBIN, J., y WATSON, G. S., «Testing for Serial Correlation, in Least Squares Regression, II», *Biometrika*, 38, 1951, 159-178.
15. DURBIN, J., y WATSON, G. S., «Testing for Serial Correlation in Least Squares Regression, III», *Biometrika*, 58, 1971, 1-19.
16. FRIEDMAN, M., y MEISELMAN, D., «The Relative Stability of Monetary Velocity and the Investment Multiplier in the U.S. 1897-1958», en *Comision on Money and Credit, Stabilization Policies*, Englewood Cliffs, Prentice Hall, Nueva York, 1963.
17. GOLDFELD, S. M., y QUANDT, R. E., *Nonlinear Methods in Econometrics*, North Holland Publishing Company, Amsterdam, 1972.
18. GRILICHES, Z., y RAO, P., «Small Sample Properties of Several two-stage Regression Methods in the Context of Autocorrelated Errors», *J.A.S.A.*, 64, 1969, 253-272.
19. HANNAN, E. J., «Testing for Serial Correlation in Least Squares Regression», *Biometrika*, 44, 1957, 57-66.
20. HANNAN, E. J., *Times Series Analysis*, Methuen and Company, Ltd., Nueva York, 1960.
21. HART, B. I., «Tabulation of Probabilities for the Ratio of the Mean Square Successive Difference to the Variance», *Annals of Mathematical Statistics*, 13, 1942, 207-214.
22. HARVEY, A. C., y PHILLIPS, G. D. A., «A Comparison of the Power of Some Tests for Heteroskedasticity in the General Linear Model», *Journal of Econometrics*, 2, 1974, 307-316.
23. HEDAYAT, A., y ROBSON, D. S., «Independent Stepwise Residuals for Testing homoscedasticity», *J.A.S.A.*, 65, 1970, 1573-1581.
24. HENSHAW, R. C., «Testing Single Equation Least Squares Regression Models for Autocorrelated Disturbances», *Econometrica*, 34, 1966, 646-660.
25. IMHOFF, P. J., «Computing the Distribution of Quadratic Forms in Normal Variables», *Biometrika*, 48, 1961, 419-426.

26. JOHNSTON, J., *Econometric Methods*, 2.<sup>a</sup> ed., McGraw Hill, Nueva York, 1972.
27. KADIYALA, K. R., «A Transformation Used to Circumvent the Problem of Autocorrelation», *Econometrica*, 36, 1968, 93-96.
28. KMENTA, J., *Elements of Econometrics*, McMillan Co., Nueva York, 1971.
29. KMENTA, J., y GILBERT, R. F., «Estimation of Seemingly Unrelated Regressions with Autocorrelated Disturbances», *Econometric Workshop Paper n.º 6805*, Michigan State University, 1968.
30. KOERTS, J., y ABRAHAMSE, A. P. J., *On the Theory and Application of the General Linear Model*, Rotterdam University Press, 1971.
31. KOERTS, J., «Some Further Notes on Disturbance Estimates in Regression Analysis», *J.A.S.A.*, 62, 1967, 169-183.
32. LOZANO, V., «Modelos con autocorrelación en las perturbaciones. Método de los mínimos cuadrados generalizados», *Estadística Española*, 27, 1965, 5-37.
33. NERLOVE, M., «Returns to Scale in Electricity Supply», en *Measurement in Economics: Studies in Memory of Yehuda Grundfeld*, cap. 7, Stanford University Press, 1963.
34. NEUMANN, J. von, «Distribution of the Ratio of the Mean Square Successive Difference to the Variance», *Annals of Mathematical Statistics*, 12, 1941, 367-395.
35. NEUMANN, J. von, «A Further Remark Concerning the Distribution of the Ratio of the Mean Square Successive Difference to the Variance», *Annals of Mathematical Statistics*, 13, 1942, 86-88.
36. ORCUTT, G. H., «A Study of the Autoregressive Nature of the Time Series Used for Timbergen's Model of the Economic System of the U.S.», *Journal Royal Statistical Society*, 10, 1948, 1-53.
37. PAN JEL-JIAN, *Distributions of the Noncircular Serial Correlation coefficients. Selected translations in mathematical statistics and probability*, vol. 7, Institute of Mathematical Statistics by the American Mathematical Society, 1968.
38. PENA, J. B., «Consideraciones Econométricas sobre Autocorrelación», *Estadística Española*, 26, 1965, 5-24.
39. PHILLIPS, G. D. A., y HARVEY, A. C., «A Simple Test for Correlation in Regression Analysis», presentado al Congreso Europeo de Econometría, Oslo, 1973.
40. SIMS, C. A., «Distributed Lags», en *Frontiers of Quantitative Economics*, II, Intritigator y Kendrick editores, North Holland Pub. Amsterdam, 1974.
41. THEIL, H., y NAGAR, A. K., «Testing the Independence of Regression Disturbances», *J.A.S.A.*, 56, 1961, 793-806.
42. THEIL, H., *Principles of Econometrics*, John Wiley Sons, Inc., Nueva York, 1971.
43. THEIL, H., «The Analysis of Disturbances in Regression Analysis», *J.A.S.A.*, 60, 1965, 1.067-1.079.
44. WOLD, H., «On Least Square Regression with Autocorrelated Variables and Residuals», *Bulletin de L'Institut International de Statistique*, 32, 1950.